Award Number: W81XWH-13-1-0093

TITLE: Genetic and Functional Heterogeneity of Tumors in Neurofibromatosis 2

PRINCIPAL INVESTIGATOR: James F. Gusella, Ph.D.

CONTRACTING ORGANIZATION: Massachusetts General Hospital
Boston, MA 02114

REPORT DATE: May 2015

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
May 2015	Annual	04/15/2014- 04/14/2015

4. TITLE AND SUBTITLE
Genetic and Functional Heterogeneity of Tumors in Neurofibromatosis 2

5a. CONTRACT NUMBER
W81XWH-13-1-0093

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
James F. Gusella, Ph.D.

E-Mail: Gusella@helix.mgh.harvard.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Massachusetts General Hospital
185 Cambridge St, CPZN 5
Boston MA 02114

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The intent of this project is to integrate whole exome genetic and global expression data to identify genes that contribute to the formation, progression and heterogeneity of NF2-associated tumors. In year 1 we prepared and submitted for exome sequencing 126 samples representing paired human tumor (meningioma or schwannoma) and normal DNAs from the same individuals. We also prepared RNA from the same tumors for transcriptome sequencing. In year 2 we completed and began analysis of these datasets as well as RNA expression of a panel of isogenic arachnoidal cell lines either heterozygous or homozygous for inactivating *NF2* mutation. These analyses have identified somatic mutations that implicate several candidate genes and cell pathways altered by *NF2* mutation. The integration of these datasets is now underway to finalize and confirm genes and processes that contribute to NF2 tumor formation and assess their effects on cellular phenotypes.

15. SUBJECT TERMS
Neurofibromatosis 2, meningioma, schwannoma, exome, transcriptome

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	UU	9	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

1. Introduction

In neurofibromatosis 2 (NF2), tumor formation requires inheritance of a germline mutation in the *NF2* gene that inactivates one allelic dose of the merlin tumor suppressor, followed by somatic loss or inactivation of the remaining allele. However, it is not known whether merlin deficiency alone is sufficient for tumor formation, either in schwannomas or in meningiomas, the two major tumor types of NF2. A number of lines of evidence suggest that other genetic lesions may participate in the initiation of NF2 tumors and that other genes certainly contribute to the functional heterogeneity and progression observed in meningiomas. In the latter instance, there are many studies that have noted rearrangements of particular chromosomal regions in merlin-deficient meningiomas, but despite many years of investigation of these regions, the critical genes that participate in tumor development and heterogeneity and the biological pathways that they represent remain unknown. We postulate that there are somatic genetic events that occur in individual genes that contribute to the formation, progression and heterogeneity of tumors in NF2 and that integrated analysis of whole exome sequence data together with tumor characteristics, genomic rearrangements, RNA expression data and microRNA expression data from these tumors in comparison with normal arachnoidal cells will reveal the biological pathways that underlie the development and functional heterogeneity of NF2 meningiomas.

2. Keywords

Neurofibromatosis 2, meningioma, schwannoma, exome, transcriptome, arachnoid, CRISPR

3. Overall Project Summary

Task 1- DNA sequence analysis of meningiomas

Our first specific aim was to perform paired whole exome sequencing of meningiomas and normal DNAs from the same individuals to identify *de novo*, somatic alterations, from point mutations and small insertion/deletions that affect protein structure to large regions of loss-of-heterozygosity (LOH). We also indicated the possibility of complementing these analyses of meningioma with similar analyses of the other NF2-associated tumor, schwannoma. The goal in Year 1 was to generate exome data from the first 30 of these meningiomas and their matching normal DNAs. We first successfully screened a large number of tumors and selected those for exome sequencing. We then generated high quality DNA from 26 meningiomas (4 sporadic and 22 NF2-associated) with matched blood DNA. We also augmented this analysis by preparing high quality DNA from 30 vestibular schwannomas (4 NF2 and 26 sporadic) all obtained through clinically-indicated surgery and their corresponding blood DNA. In addition, we prepared DNA from 8 meningiomas, 5 vestibular schwannomas, 4 nerve samples and 1 arachnoid sample from two independent NF2 autopsies. We then submitted DNA from 126 samples (63 tumor and normal DNA pairs) to the DNA Sequencing Facility for preparation of exome capture libraries, indexing and deep sequencing using Illumina HiSeq2500 DNA Analyzers. By running this larger than anticipated number of tumors in a single batch, we avoided the potential batch effects of splitting the analysis into two batches of ~60.

In Year 2, we received and analyzed the exome data from these tumors The exome sequences were generated by Illumina 76 paired-end sequence reads at the Broad Institute Genome Analysis Platform. Sequence reads that were demultiplexed by Illumina's CASAVA software, were mapped to the human genome build 37 (hg19) through BWA v. 0.5.9 [1] with parameters -q 5 -l 32 -k 2 –o 1. The resulting alignments were further sorted, indexed and duplicates removed using the Picard Tools [http://broadinstitute.github.io/picard/]. Next, base quality score recalibration and local realignment around known indels were performed using GATK [2,3]. In this process, five meningioma and two schwannoma samples failed to pass the QC metrics and thus, they were removed from more detailed analysis.

Somatic variants were called applying muTect v.1.1.4 [4] with parameters –min_qscore 20 –clipping_bias_pvalue_threshold 0.05 and Indelocator [https://www.broadinstitute.org/cancer/cga/indelocator] with a parameter –ws 300 to blood-matched tumor pairs to identify single nucleotide variants and small indels respectively. Variants were identified only in the target exon intervals used in the exome capture sequencing. Next, identified variants with the respective read depths ≥8X and ≥14X in paired blood and tumor samples, and tumor fraction ≥ 0.1 were annotated using Oncotator v.1.3 [5] with the oncotator_v1_ds_Sept172014 database.

Additionally, alignability scores (≥ 0.8) based on alignability of 36mers [6] were applied to further filter the variants.

We also performed dosage analysis using the exome read counts to define chromosomal rearrangements in the tumors. First, we computed read depth coverage metrics for target intervals of at least 20 bp in length that were used in the exome sequencing for each sample in the analysis using GATK's DepthOfCoverage walker with –minMappingQuality 20 option [2]. Further, coverage metrics were GC-content normalized as described in [7], where GC-content for target intervals was calculated by GATK's GCContentByInterval walker [2]. Next, log2 ratios of normalized read depth between tumor and matched blood for each target interval with > 5X coverage in both test and control samples were calculated. Log2 ratios were further scaled by subtracting the median of all log2 ratios in the sample. Relying on scaled log2 ratios, we used R's DNAcopy package to segment the data. Later, segments were merged with VarScan's mergesegments.pl script using -0.4 and 0.4 thresholds for deletions and amplifications respectively [8].

Consistent with prior studies from ourselves and others using dosage array analyses, the exome read counts suggested extensive regions of chromosome loss and some regions of duplication, with a much higher frequency of such events in meningiomas than in schwannomas. These events have been catalogued by tumor and we are currently merging the calls with loss-of-heterozygosity calls from the variant analysis above to obtain the most accurate assessment of the extent of *de novo* deletions in the tumors.

For defining candidate genes whose inactivation contributes to tumorigenesis, we identified all *de novo* somatic variants in the exome data and classified them into loss-of-function mutations (nonsense, frame shift deletion and insertion, and splice_site), missense mutations with deleterious effect predicted by PolyPhen [9], Sift[10], Mutation Assessor [11], Mutation Taster[12] and LRT [13], and other non-synonymous mutations such as in_frame indels, and enforcing ≥10X coverage in blood samples. There was no single gene mutated in a high proportion of the tumors other than *NF2*. However, a small subset of genes each displayed two independent somatic hits across these tumors, making these loci candidates for further analysis. These genes, listed in Table 1, were mutated primarily in schwannomas rather than in meningiomas, which, along with the difference in the frequency of dosage change, suggests a different genetic architecture in these two tumor types, with the exception of chromosome 22 loss (the site of the *NF2* gene) which occurs at very high frequency in both. We are currently performing Sanger sequencing confirmation of each of these mutations prior to deciding on whether to perform targeted sequencing of additional tumors for the genes that qualify. We are also awaiting final dosage/loss of heterozygosity analysis to identify any single hits from the exome analysis of either tumor type that would be added to the candidate list because they align with regions of deletion and to identify any regions lost in the meningiomas that affect the same genes affected by point mutation in schwannomas.

Table 1- Candidate contributors to tumorigenesis based upon recurrent mutation

Gene Name	Location	No. Exons	Total gene size (bp)	Number of Tumors	cDNA change	protein change
Schwannoma						
PAK2	3:196466728-196559518	15	92791	2	c.1493C>t,c.383A>G	p.P498L,p.K128R
PRDM1	6:106534195-106557814	7	23620	2	c.2119G>A,c.202delA	p.E707K,p.K68fs
NUP153	6:17615266-17707065	22	91800	2	c.1633A>G,c.3897_3899delTGC	p.I545V,p.A1300del
RYR2	1:237205702-237997288	105	791587	2	c.2080C>T,c.10820_10821insT	p.R694*,p.PL3607fs
UNC5A	5:176237560-176307899	15	70340	2	c.1103C>T,c.1220G>A	p.P368L,p.R407H
ZCCHC14	16:87439852-87525460	13	85609	2	c.25A>T,c.17C<G	p.S9C,p.P6R
Meningioma						
NLRP1	17:5404719-5487832	16	83114	2	c.1853delG,c.916G>C	p.S618fs,p.E306Q

Task 2. RNA and microRNA sequence analysis of meningiomas

Our second specific aim was to perform RNA sequencing (RNA-seq) to define both gene and microRNA expression profiles relative to normal arachnoid tissue in the same set of meningiomas used for exome analysis above. In Year 2, we performed RNAseq and miRNAseq for 23 primary meningiomas obtained at surgery, 4 meningiomas obtained at autopsy, 4 arachnoid tissue samples obtained at autopsy, a cultured benign meningioma line (BenMen1), a *NF2* heterozygote arachnoidal line (AC7_A3) and 3 *NF2*-null arachnoidal lines isogenic with AC7_A3 except for CRISPR/Cas9 editing of *NF2* (see Table 3). RNAseq libraries were prepared using a customized version of a strand-specific dUTP method [14,15]. Libraries were sequenced with 76-bp paired-reads on the Illumina platform. The quality of raw sequence reads were assessed by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and they were further quality trimmed using sickle with options –q 20 –l 70 [16]. All autopsy tissue and 1 surgical meningioma specimen libraries failed quality control metrics and were not analyzed further. Sequences for the remaining 27 libraries were aligned to the human genome (GrCH37, Ensembl build 71) using Gsnap [17] version 2014_12_19. Expression levels of genes in the units of count-per-million were estimated by using bedtools and Ensembl's gene annotation as described in [18] based on uniquely aligned reads. Figure 1 shows a multi-dimensional scaling plot (equivalent to Principal Component Analysis) comparing gene expression in these samples, revealing close clustering of all 3 CRISPR/Cas9 edited *NF2*-null lines and the BenMen1 line, with clear separation from the wild-type arachnoidal line, AC7_A3. Also evident is a spread of the primary meningioma samples suggesting greater heterogeneity of expression than the cultured lines, consistent with the extensive dosage changes seen in Task 1.

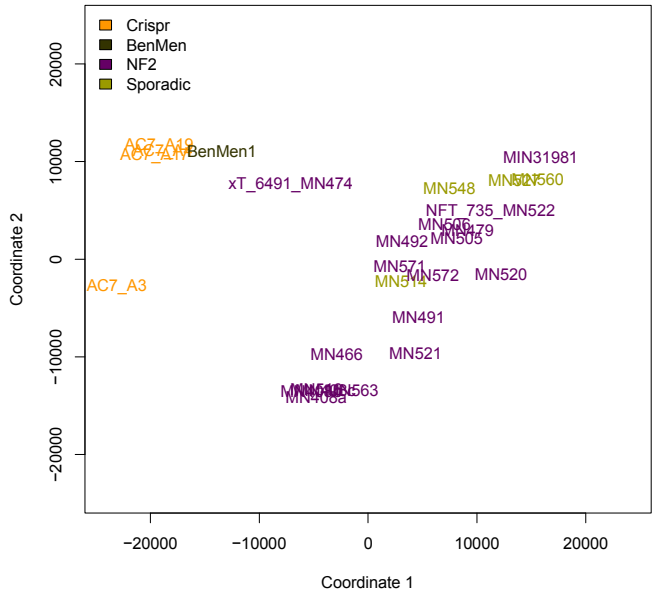

Figure 1- MDS Plot based upon global gene expression

For an initial analysis of differential expression analysis, we calculated log2 ratios of expression values between test samples and controls and further converted them to z-scores. To identify the statistically significant genes expressed differentially between AC7_A3 and the three *NF2*-null arachnoidal lines, we chose those genes with absolute z-score values in top 5%. When analyzing all tumor samples, we compared to AC7_A3 and converted z-scores to p-values under the normal distribution and selected genes with p-value < 0.05 as differentially expressed genes. We then performed GO term and pathway enrichments by DAVID [19] relying on differentially expressed genes identified in the analyses. These analyses remain preliminary as they have not yet been integrated with the mutation and dosage analysis of Task 1, which is a goal of Task 3 in Year 3. However, two features are already evident from the pathway enrichment studies. Both the *NF2*-null lines and the meningiomas show an enrichment among down-regulated genes for terms related to metal ion and cation binding while the meningiomas show a strong enrichment among "upregulated" genes for terms related to immune processes, suggesting the presence of infiltrating immune cells in the tumors. One of the difficulties with more detailed analysis is the lack of the normal arachnoidal tissue comparator. We plan to

make a more detailed analysis of the data after performing RNAseq on many more replicates of the cultured arachnoidal cells to achieve more reliable statistics.

In Year 2, we also prepared and sequenced 36 miRNA libraries 36bp single-end reads on the Illumina platform. The average number of reads per library was 31.7 million and one library failed in sequencing. The raw reads were first trimmed against small-rna sequencing adapters using cutadapt v. with -e 0.1 –O 5 –m 5 options [20]. We next aligned the reads with length between 16 and 25 nt to known mature miRNA sequences from miRBase database [21](release 21) using BWA aln with –n 1 option [22]. The average number of mapped reads per library was 4.56 million as seven libraries yielded less than two million reads. We have not yet performed more extensive computational analysis of these libraries as we plan to rerun them to obtain more reads, but will perform the final analysis as part of the integration in Task 3.

Task 3. Identification and validation of genes contributing to tumorigenesis

This task comprises specific aims 3 and 4 and was slated to begin only in Year 3. Our third specific aim, to use integrated analysis to identify the genes/pathways that are implicated by somatic alterations as cooperating in tumor formation or progression to a higher grade tumor, actually began late in Year 2 when the first batches of data from the exome and RNA sequencing both became available. There are no definitive results to report as yet, but the analysis is aided by our ability to compare the RNA sequencing results of tumors with the baseline effects on global gene expression in *NF2*-null arachnoidal cells that we created in Year 1 and are listed in Table 3. These cells are also invaluable for our fourth specific aim, which is to test whether changes associated with alterations of tumor-associated genes in Aim 3 are reproduced by their knock-down in normal and merlin-suppressed arachnoidal cells. We had planned to use RNA interference, specifically employing lentiviral-delivered short-hairpin RNAs (shRNA), to specifically suppress the expression of target genes, but our experience to date with CRISPR/Cas genome editing indicates that introducing targeted inactivating mutations will be a much cleaner and more informative approach.

Table 3 – Isogenic Arachnoidal Cells

Cell line	clone #	NF2 expression	*NF2* mutation genotype
AC_7	A3	+/-	Exon 15, c.1599_1602delGCAT (het) Exon 8, wildtype
AC_7	A4	-/-	Exon 15, c.1599_1602delGCAT (het) Exon 8, c. 787del23bp (het) Exon 8, c. 804insC (het)
AC_7	A17	-/-	Exon 15, c.1599_1602delGCAT (het) Exon 8, c. 795del8bp (het) Exon 8, c. 802insT (het)
AC_7	A19	-/-	Exon 15, c.1599_1602delGCAT (het) Exon 8, c. 787del35bp (hom)

4. Key Research Accomplishments

- Completion and analysis of exome sequencing of 126 samples representing human tumor (meningioma or schwannoma) and normal DNAs from the same individuals

- Generation and preliminary analysis of RNAseq data from the meningiomas used for exome sequencing

- Generation and preliminary analysis of RNAseq data from isogenic arachnoidal cells with or without inactivation of *NF2*, generated using CRISPR/Cas genome editing technology

5. Conclusion

The goal of Year 3, that of integrating genetic and gene expression data from NF2-associated tumors is now possible due to our generation of exome and RNA sequencing data. In addition, our generation of RNA sequencing data from the novel isogenic arachnoidal cell system created using CRISPR/Cas genome editing is invaluable for guiding this integration by providing a clean baseline of the effects of the *NF2* mutation alone. The cells will also permit us to introduce targeted mutations for top candidate tumorigenesis genes, several of which are currently being confirmed, and to determine the effect on gene expression and cellular phenotypes for a more complete understanding of NF2 tumorigenesis.

6. Publications, abstracts and presentations

None yet

7. Inventions, patents and Licenses

None

8. Reportable Outcomes

The results are yet to be reported.

9. Other Achievements

None

10. References

[1] Li H., Durbin R. Bioinformatics 2010 Mar 1;26(5):589-95.

[2] McKenna A. et al. Genome Research 2010 Sep; 20(9):1297-303.

[3] Geraldine A. et al. Current Protocols in Bioinformatics 2013 Oct 15;11(1110):11.10.1-11.10.33.

[4] Cibulskis K. et al. Nature Biotechnology 2013 Mar; 31(3):213-9.

[5] Ramos A.H. et al. Human Mutation 2015 Apr; 36(4):E2423-9.

[6] Darrien T. et al. PLoS One 2012;7(1):e30377.

[7] Magi A. et al. Genome Biology 2013, 14:R120

[8] Koboldt DC. et al Genome research 2012, 22(3):568-76

[9] Adzhubei I.A. et al. Nature Methods 2010; 7(4):248-9.

[10] Kumor P. et al. Nature Protocol 2009; 4(7):1073-81.

[11] Reva B. et al. Nucleic Acids Research 2011;39(17):e118.

[12] Schwarz J.M. et al. Nature Methods 2010; 7:575-76.

[13] Chun S. et al. Genome Research 2009; 19:1553-61.

[14] Levin JZ et al. Nature Methods 2010; 7(9):709-15

[15] Parkhomchuk D et al. Nucleic Acids Research 2009; 37(18):e123

[16] Joshi NA, Fass JN. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]

[17] Wu TD et al, Bioinformatics 2010; 26(7):873-81

[18] Blumenthal I, et al. AJHG 2014; 94(6):870-83.

[19] Huang D.W. et al. Nature Protocols 2009; 4:44-57.

[20] Martin M. EMBnet.journal 2011; 17:10-12.

[21] Kozomara A., Griffiths-Jones S. 2014; 42:D68-73.

[22] Li H., Durbin R. Bioinformatics 2009; 25:1754-60.

11. Appendices

None